Best Practices Guide for Conducting Assessments in Counterinsurgencies

by Dave LaRivee

Purpose

This guide provides practical advice to assessment strategy planners and practitioners. It aims to fill the gap between instructions provided in handbooks and field manuals, and the challenges faced when adapting these instructions to specific operations.   Its purpose is to complement, not replace, the more detailed planning or instructional documents. Wherever possible, the articles in this guide provide references to more detailed assessment planning documents. It also makes reference to some of the specific needs of the implementation of the Transition (Inteqal) process in Afghanistan.

Introduction

Assessments are difficult to conduct even under the best of conditions. In practice, the assessment's goal is to examine the most recent states of a diverse set of conditions in order to measure the contribution of multiple and often countervailing actions on progress towards a broadly defined set of objectives. This would be hard enough in a stable or mildly dynamic environment. Conducting assessments in the middle of a counterinsurgency campaign, however, introduces a host of additional challenges.   For one, the countervailing actions are no longer accidental; they are the deliberate attempts by insurgent forces to negate the efforts of the counterinsurgent force.   Additionally, the relationships between an action in one dimension and effects in another are often poorly understood or dependent on a potentially endless combination of initial conditions.

In this light, it is difficult to prescribe a fixed set of procedures and guidelines that fully prepare assessment teams for the challenges of assessing counterinsurgency campaigns. Most currently available sources provide definitions of key terms, outline the main processes involved, and explain how assessments support operations. But these handbooks and manuals do not fully prepare the practitioner for the messy conditions and shifting demands of real-time assessments. There is an enormous gap between how we are taught to conduct assessments and how we actually conduct assessments.

Fortunately, we have learned much about assessments from recent practical experience. This guide attempts to close the gap between the ideal and the reality of assessment by providing insights into the "philosophy" of assessment, highlighting the challenges, and sharing best practices from the field used to address these challenges.

To make this guide immediately useful to the practitioner, its recommendations assume the continued existence of major structural obstacles to an accurate, transparent, and credible assessment and offers suggestions for working around these obstacles to minimize their negative

August 17, 2011

Report Documentation Page		*Form Approved* *OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **17 AUG 2011**	2. REPORT TYPE	3. DATES COVERED

4. TITLE AND SUBTITLE **Best Practices Guide for Conducting Assessments in Counterinsurgencies**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **United States Air Force Academy,Colorado Springs,CO,80840**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**35**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

consequences.   This guide shows how to do an assessment today, not how to change the future assessment environment.

Approach

The guide includes twenty articles that address the most prominent issues assessment teams face in the field. The articles broadly address the assessment philosophy (Part One) or assessment method (Part Two).

The assessment philosophy articles seek to clarify assessment's purpose and objectives. By reminding practitioners of how their assessments can be used to influence the overall campaign strategy, it becomes easier to make the right choices between sources and methods. It also helps practitioners understand how to build and communicate an assessment that will influence strategic decisions.

The method-oriented articles are more tactical in nature. These methods will be familiar to most practitioners and do not include many examples of groundbreaking innovations. But knowledge of proper methods is no guarantee of their effective application. In practice, we often do not remember to use these methods, or do not apply them in a creative fashion. In the demanding, complex, and time-sensitive world of assessment, we rush to deliver a product, but may not realize that we are not delivering the right product. Part Two highlights some of the most common assessment pitfalls, reminds us of some fundamentals, and offers creative means for dealing with intransigent players or intractable obstacles. An overview of the twenty articles is offered below.

Overview

Part One: Assessment Philosophy

Article One: Remain True to the Assessment's Objective. The objective of an assessment is to produce insights pertaining to the current situation, and to provide feedback that improves the decision maker's decisions. This article discusses how key elements of this objective should guide the assessment development process.

Article Two: Take a Multi-dimensional Perspective. This article describes why it is essential to build the assessment by looking at the environment through multiple perspectives that cross lines of operations and time periods. It also highlights some errors that may arise if the assessment lacks a broad perspective.

Article Three: Serve as the Bodyguards of Truth. Assessment teams develop what may, by default, become the only publicly-available, official picture of the campaign. Therefore, assessment teams must serve as the bodyguard of truth and never compromise the integrity of their reports. This article outlines nine key practices that help preserve the integrity of assessments.

Article Four: Ensure Independence and Access. Strategic assessment teams need to be free to express their findings about the current conditions and the influential factors they discover. They also need access to a wide array of information and people in order to perform their job properly.   This article describes how to secure independence and access through a partnership between the senior sponsor of the assessment team, individual line of operation owners, and the assessment team.

Article Five: Nurture the Intelligence - Assessment Partnership. The activities related to intelligence and assessments often seem remarkably similar, thus generating the potential for confusion or duplication of effort. This article briefly discusses the mutually supporting relationship between the two activities. It uses references from formal documents and recommends that the leaders of the two communities deliberately develop a shared understanding of this symbiotic relationship in order to avoid problems.

Part Two: Method

Article Six: Establish a Terms of Reference Document. Unclear terms generate confusion in the design of the assessment framework, the analysis of data, and the reporting of insights. Thus, it is in the team's best interests to develop a Terms of Reference document as soon as possible.

Article Seven: Build the Assessment Framework Iteratively, Incrementally, and Interactively. The assessment framework should be built in stages through a collaborative process. This approach minimizes complexity, allows for effective learning, and retains clearly established priorities. It also allows the assessment team to refine the focus and scope of the assessment framework based on lessons learned during the development and use of earlier versions.

Article Eight: Discriminate between Indicators and Metrics. Most people use the term indicator and metric interchangeably and suffer little or no consequences or confusion. However, there are times when it is useful to discriminate between the two. This article offers a useful approach for when and how to discriminate.

Article Nine: Use Each Class of Indicator Properly. Some indicators can be grouped into classes because they share a common set of characteristics that may be beneficial or detrimental to the assessment process. Several of these broad classes are described in this article including those that measure input versus outcome, those that indicate failure to achieve a condition (spoilers), metrics that can indicate positive or negative effects depending upon context (bipolar), and those that serve as substitutes for other hard-to-measure indicators (proxies).

Article Ten: Beware of Manipulated Metrics. Some metrics can be manipulated by the subjects under observation to send misleading signals to observers, rather than reflecting the reality of the current conditions. This is a particularly high risk for metrics that are used to promote or demote, or directly redistribute resources and money.   This article discusses several examples and suggests ways to detect and minimize such distortions of the data.

Article Eleven: Develop a Manageable Set of Metrics. There are hundreds of metrics available for consideration at any point in time. Thus, it is necessary to establish rules that help us select the metrics contributing the most to the assessment effort. This article discusses several screening filters that help practitioners develop a manageable and effective set of metrics.

Article Twelve: Retain Balance in Both Metrics and Method. Interrelated debates arguing the merits of the narrative versus summary graphics, the organizational level at which assessments should be performed, and the need to preserve the front-line commander's views within higher level summary assessment products persist in the assessment world. This article

suggests using a format that balances different metrics and method to capture the best features of each alternative.

Article Thirteen: Deploy Field Assessment Teams. In order to provide actionable information to the decision maker, assessment insights must be relevant and credible. For critical issues, the only way to achieve this standard is get out to the field and engage directly with front-line units. This article suggests that we rethink how we perform assessments and offers an approach that augments the traditional process with the use of field assessment teams.

Article Fourteen: Bound Estimates with Eclectic Marginal Analysis. When a desired metric is difficult to measure directly we might be able to measure other factors that drive the value of the desired metric. Under such conditions, we can use marginal analysis with an eclectic set of related metrics to generate a reasonable estimate of the target metric. This section explains the technique and provides some examples of marginal analysis.

Article Fifteen: Anchor Subjectivity. A degree of subjectivity in assessments is unavoidable. This article discusses methods to minimize the degree of subjectivity, make that subjectivity transparent, and maintain consistency in the way we capture subjective assessments.

Article Sixteen: Share Data. Every coalition effort faces information sharing challenges. This article discusses important reasons for sharing information and offers some guidelines that promote effective sharing.

Article Seventeen: Include Host Nation Data. Two features of the COIN assessment environment that should be considered when developing the assessment process are the existence of host nation data collection efforts and the ability for assessment teams to interact with this system. This article addresses the challenges of using host nation data and ways to work around the challenges.

Article Eighteen: Develop Metric Thresholds Properly. This article discusses key guidelines for developing metrics thresholds, including adjusting levels towards key phases of objective conditions, developing and sharing clear definitions of the thresholds, and ensuring that observances of metrics at these levels represent a significant change in underlying conditions.

Article Nineteen: Avoid Substituting Anecdotes for Analysis. Anecdotes are a useful component of assessments when used properly. Unfortunately, they are often used as substitutes for a solid assessment. The best rule to keep in mind when using anecdotes is that they are generally the starting point for analysis, not the closing argument of an assessment.

Article Twenty: Use Survey Data Effectively. Questions of motivation, satisfaction, degrees of trust or fear, as well as intentions regarding future actions are difficult to measure by monitoring actions. Often, we must capture this information by interviews or broader surveys. This article addresses how to manage some of the major concerns associated with using survey data in assessments.

Part One: Assessment Philosophy

Article One: Remain True to the Assessment's Objective

However beautiful the strategy, you should occasionally look at the results --Churchill

A clear definition of the assessment objective helps us define the process and products to properly support the desired campaign objective. Our working definition of the assessment objective is to produce insights pertaining to the current situation and progress of the campaign, thereby providing feedback that improves the senior leader decisions. The key parts of this objective that guide the assessment development process are discussed below.

First, the assessment must synthesize insights gained from the available information. This requires providing the insight behind the numbers at each stage of the process instead of forwarding numbers up the chain without context. The accompanying context presents a richer picture of the situation and helps reviewers at higher levels decide when it is no longer appropriate to aggregate data. In one example of a flawed metric, unit reports on the number of districts stabilized were increasingly aggregated beyond the point where the meaning was clear. However, if reports focus on insights, it will be clear to the higher reporting level that while 70% of the districts are stable, the unstable districts account for 80% of the population. Thus, in this case, the aggregate metric of stable districts had lost its value as an ordinal measure of progress

Second, an assessment should provide a feedback on our efforts. In the case of the transition process in Afghanistan (Inteqal), the Joint Afghan-NATO Inteqal Board (JANIB) assessment must evaluate the readiness of individual provinces or districts for transition of their security to the government of Afghanistan across all the functions of government. Throughout the multi-year Inteqal process, we are looking to close the gap between current conditions and the necessary initial conditions for transition. Not only must assessors identify these gaps, they must explore the root causes of the gaps and examine the potential remedies, often based on what worked in other provinces. If metrics and assessment reports do not provide this feedback, or we cannot draw useful insights for closing the gap, then reporting is at best incomplete. Eliminating metrics that fail to meet these requirements should reduce the number of metrics reviewed each reporting period since many commonly used metrics provide no feedback on the effectiveness of COIN activity and many provide no insights on conditions or progress (NATO Handbook, 8.1.2 MOE Considerations).

Finally, the insights need to be linked to a particular course of action available to the decision maker and they must be sufficiently supported to justify a decision or action. In his January 2010 report, Major General Flynn criticized the intelligence field as "a culture that is strangely oblivious of how little its analytical products, as they now exist, actually influence commanders" (Flynn). As full partners with the intelligence community (see Article Five), assessment teams are subject to the same criticism. The key point here is that assessment teams often fail to recognize this clear indicator of our own effectiveness—a successful assessment is the one that influences the customer. Developing assessments that are relevant and actionable is an iterative process. Subsequent articles describe several ways to make assessments more influential with decision makers.

Providing assessments that can influence the appropriate course of action will be critical to the Inteqal process. The JANIB is currently reviewing each district to determine if they meet the initial conditions for the start of transition. When districts do not meet these conditions, the Afghan government and its international partners must develop Action-Plans outlining actions required to meet these initial conditions. But even in the cases of districts that have met the conditions for **starting** the transition process, in order to meet the conditions for **completion** of transition, the Afghan government must be able to maintain these conditions independently (Joint Framework for Inteqal, p. 23). Thus, in both the assessment and implementation phases of

Inteqal, it is critical that the assessment process provide actionable insights to help the leadership develop Action-Plans. This involves prioritizing data collection and analysis to develop and assess courses of action tied to the Action-Plans. We must avoid the mere reporting of a standard set of metrics. We must continually refine our assessment focus to meet the needs of the transition process.

Article Two: Take a Multi-dimensional Perspective

The government's one-dimensional conception of a multi-dimensional process ensured its defeat
—Jeffrey Race*, War Comes to Long An*

Every COIN campaign unfolds in multiple dimensions (e.g. political, diplomatic, economic, rule of law, security) at varying levels in different districts over various time periods. Thus, we can only understand current conditions or progress made, and recommend corrective action by looking at indicators of conditions through multiple filters that capture these dimensions. While we may be accustomed to view sets of indicators as belonging to unique dimension, this approach can often lead to an improper reading of the situation. Taking a multi-dimensional perspective helps avoid such errors, but to be successful the assessment team must understand the interrelationships between effects and their indicators across multiple dimensions.

A multi-dimensional perspective also helps better understand the enemy. COIN is a multi-stage learning contest where both sides use what they learn in one period to adapt to each other's moves. To understand how insurgents adapt we need to look across dimensions and over time to capture both the diverse direct effect on conditions and the multiple ways in which insurgents respond to our actions. For example, when we act in the security dimension insurgents may respond in the economic or governance arena. We may find that what looks like an ineffective action in the security dimension, may have generated a significant effect in another dimension.

Assessing the delivery of essential services illustrates the challenges and gains from a multi-dimensional perspective. A rudimentary assessment of essential services may look at the hours of electrical power provided within a district to evaluate the government's ability to deliver essential services. In general, more hours is considered better. However, since the purpose of the campaign is to facilitate improvement in the delivery of essential services to reinforce the government's legitimacy, the assessment goal is not just to report status. If the hours of power are lower than the established benchmark, we need to know the reason for the shortfall and we should look across all possible dimensions of the issue. The problems could be rooted in poor governance (a lack of capacity in budgeting or planning), a struggling economy (generators lack parts or fuel supplies), weak security (transmission lines are attacked, fuel deliveries are hijacked), or the absence of the rule of law (corrupt officials steer power across the grid), or even diplomatic failure (if power grids are internationally linked, neighbor states could be rationing access due to lack of support for the nation's government or a dispute over refugee flows).

It is obvious that a failure to look at this problem through each dimensional lens risks failing to identify the principal source of the problem and the best means for resolution. Another advantage of taking a broad perspective is that it identifies multiple means for solving the problem. Achieving minor, but synergistic, gains in several dimensions could collectively generate enough pressure to overcome the problem.

A second example illustrates another benefit of a multidimensional perspective.   One of the key objectives of security-oriented actions is to create enough breathing space to allow the pursuit of other activities such as negotiations to reconcile warring parties, completion of infrastructure projects, or reopening businesses and banks.   When we look at our indicators in the security dimension we may be assessing them against some prior level of security incidents and note significant gains from previous levels. However, if the gains are insufficient to allow breathing space for the aforementioned activities along other lines of operation, then the gains are really insignificant relevant to their main purpose. This example suggests that even the thresholds we set for indicators in one dimension are dependent upon effects in other dimensions. Recognizing this reality allows us to create more informative metric sets and their associated thresholds.

Article Three: Serve as the Bodyguards of Truth

In wartime, truth is so precious that she should always be attended by a bodyguard of lies
— Churchill

During WWII Winston Churchill insisted that the plans for the Normandy invasion be hidden behind a bodyguard of lies. But, the nature of warfare has since changed significantly, particularly in counterinsurgency operations, where deception can work against the COIN strategy. To assure the host nation's population of the legitimacy of their government, to retain the support of our coalition's governments, and to partner with the media productively we must build trust. And, only truth can build trust. The demands for transparency and credibility require that we present our assessment of progress in a fair and accurate manner. As the authors of what often becomes the publicly-available, official picture of the campaign, the assessment team must serve as the bodyguards of truth and never compromise the integrity of their reports. The analyst draws his influence and power to persuade from his analytical independence and integrity. These two qualities must be carefully guarded.

There are a few key practices that support the practitioners' role as bodyguards of truth:

(1) Don't make up data that you don't have. If requested information is not available, highlight the gap and try to adjust your collection process to get what you need using proxy data (see Article Nine: Classes of Indicators). In recent reviews of the assessment process some field analysts admit to making up data that was unavailable or too difficult to collect. As a recipient of the data, I would rather not get anything, than unknowingly build an assessment on manufactured data.

(2) Data collection and analysis teams will find it hard to comply with the "don't make it up" rule if commanders at the top are unwilling to accept no for an answer.   Commanders must understand this pressure and balance their requests against the reality of the security environment, availability of data, and reliability of raw information from the field. Often, the way to reconcile the commanders' demands with the lack of data is to focus on the commander's underlying question and request the field team answer the question with the best available data, rather than just reporting a specific metric. This gives the data collectors a wider set of options for meeting the commander's need.

(3) Support your findings with the information that led you to your conclusion, so your assessment remains transparent to the recipient. Most assessments are to some degree

subjective. Demonstrating the logic behind your conclusion illustrates the balance of fact and judgment in your findings and allows the user to weigh its value in his decision appropriately.

(4) Your reports must represent what you believe to be true. Don't adjust them for other opinions you don't support. Where there are sizable disagreements with others that cannot be factually resolved capture these as "views of others" so they are part of the report and remain open for discussion, but aren't taken for granted.

(5) Disagreements over assessments will arise between people of significantly different ranks and the junior member will sometimes feel deliberately pressured to adjust their assessment to accommodate the senior member. This problem arises for both junior and senior analysts. Preserve the integrity of your assessment by illustrating what improbable assumptions have to be true for your superior's assessment to be valid. Once they recognize the improbability of the required assumptions, most superiors accept the original assessment. However, your superior may remain unconvinced of the suspected flaws in his assessment and chose to present his version to others. In such cases, be clear that you still do not agree with his assessment and try to point out the public and operational risks of reporting such an assessment. You should also ensure he understands that at this point the assessment product represents his personal assessment, not that of your team, since you could no longer state in good faith that you believe it to be accurate. Working relationships between analysts and their superiors readily survive such frank and objective discussions over assessment findings.   The greater damage to a strong working partnership between analysts and their customers results from a compromise of integrity. If you lose your independence, you lose most of your value to your audience.

(6) Educate the main users and customers on your metrics and processes. The assessment community in Iraq developed a two-day workshop through which they discussed all of the major assessment tools and processes with key players in the coalition as well as five top media officials. As a result, participants were able to better understand how accuracy varied across methods and how to interpret assessment products in a way that was consistent with the underlying information.   Several of these media leaders were later instrumental in helping their media peers and audiences better understand regular assessment reports.

(7) Avoid implying greater precision than you possess. Assessment is focused on identifying broad trends and relationships. Small numerical differences are statistically insignificant due to inherent problems with the accuracy of information and reasonable burdens of proof. Small differences are even less important operationally.   If you are asked to give a number as an answer it is better to provide a range rather than a precise number.   Reacting to small numerical changes is fraught with danger. Reporting with exaggerated precision may call into question your understanding of the environment and your judgment, if not your integrity.

(8) Test the judgments of others against the hard evidence. You will receive a variety of inputs from commanders, directors, and their staffs. All of these reflect some degree of subjectivity. You must try to reduce their subjectivity with information you possess or at least evaluate that subjectivity for historical and cross-dimensional consistency.

(9) Most importantly, demand these previous eight best practices of all of your sources because once you use their information it becomes your information and your reputation as a bodyguard of truth is on the line.

Article Four: Ensure Independence and Access

Strategic assessment teams need access to a wide array of information and people in order to perform their job properly. In addition, as they develop their final product these teams need to be free to explore promising avenues of investigation. To promote these privileges in pursuit of a better product, assessment teams must have both **independence and access**. Such privileges can be secured through a partnership that includes the senior sponsor of the assessment team, individual line of operation owners or division commanders, and the assessment team. Each member of this partnership plays a critical part, outlined below.

The senior sponsor, typically the Senior Civilian Representative or Commanding General must provide a clear charter to the assessment team that achieves several objectives. First, the charter establishes the team as his lead agent for assessments. Second, it requests that all subordinate commanders designate a senior representative to support the assessment team-- typically their chief of staff or staff director. Third, it promises the subordinate commanders/directors that they remain the principal voice for assessing progress in their area or line of operations and that the assessment team will include that voice in their assessment, subject to a critical review of progress from a multi-dimensional, theater-wide perspective. Finally, it asks commanders and directors to openly share their perspectives and information with the assessment team in order to facilitate this critical review.

In return for independence and access the assessment team must fulfill a vital professional obligation. In every part of the assessment process, the team must behave as a **trusted agent** of the senior leadership throughout the chain of command. Their integrity in all actions must be above question. When operating within subordinate units they must treat all communications and information with care. Assessment products should be cleared with the host chief or director prior to their release from the team. As the assessment team develops its findings they should be shared with the host unit prior to sharing with anyone else. This "peer review" is essential to capture feedback and correct misunderstandings; but most importantly, it allows the host the first opportunity to convey new findings and possible responses up the chain of command. The assessment team's job is **not** to announce breaking news to their senior sponsor. The **"first right of disclosure"** always belongs to the host unit. The assessment team's job is to provide feedback at all levels to improve performance. The sooner the information is in the action officer's hands, the quicker things can improve. Building a team of trusted agents should not be taken lightly. The team lead needs to be chosen carefully for his or her ability to convey this trust to senior civilian and military leaders. The team lead must also actively mentor team members to ensure they preserve this trust.

In practice, disagreements occur, access is denied and units refuse to collect or release information. If the assessment team chief cannot resolve the problem with the host chief of staff, he must elevate the issue to his own chief of staff to let that person resolve the issue. But, the key to resolving most of these issues has historically proven to be the **prior** reputation of the assessment team, in the same unit or with others, as true trusted agents—that fairly represented all views, preserved the "first right of disclosure" of host commanders, and focused on supporting all insights with sound operational arguments.

Having laid out the role of the first two of the three partners, the role of the host commander/director should be fairly clear. The host recognizes the independence of the assessment team and protects their access to the host's own team. The host shares information

with the team, and in return can expect to retain the first right of disclosure. The host also recognizes the team's obligation to provide an independent critical review. The director/commander should try to shape the team's view through open discussion prior to its completion, but the host does not get to control the conclusions of the review —the host already uses the formal chain of command to provide his view and can use the same means to exercise his first right of disclosure. But the host's report does not preclude the assessment team from completing and reporting their assessment. The host enjoys a *first* right of disclosure, not an exclusive right of disclosure.

Some may view the above description as a fairy tale view of the relationship between these three key players. Historically, most Division Commanders and Line of Operation (LOO) owners are extremely receptive to this partnership once they recognized the assessment team members as trusted agents who work as much for their hosts as they do for the Senior Civilian Representative or Commanding General (CG). Unfortunately, some did reject the concept of a trusted agent outright and denied access for fear of "backdoor reports directly back to the boss". As a rule, the group that partners with assessment teams is far more likely to be effective and to face fewer unpleasant surprises than the latter group.

Article Five: Nurture the Intelligence-Assessment Partnership

Intelligence is not information alone, but also judgment— Carl Sagan

The activities of personnel involved in intelligence and assessment often seem remarkably similar and thus inevitably, the question arises regarding the difference between the two. US Army field manuals strive to discriminate between the two.   But in the end the two processes must be viewed as highly interdependent and mutually supportive.

FM 2-0 Intelligence states: "the purpose of intelligence is to provide commanders and staffs with timely, relevant, accurate, predictive, and tailored intelligence about the enemy and other aspects of the Area of Operations. Intelligence supports the planning, preparing, execution, and assessment of operations. The most important role of intelligence is to drive operations by supporting the commander's decision making." FM 2-0 further states that assessment plays an integral role in all aspects of the intelligence process.

Unfortunately, this definition is still confusing since it uses the word intelligence as part of the definition of intelligence. But the manual later discusses the components of intelligence as "detailed information, assessments, and conclusions," thus shedding some light on the difference.

FM 5-0 The Operations Process, Chapter Six, describes assessment as the continuous monitoring and evaluation of the current situation and progress of an operation. More specifically, it states that assessment deliberately compares forecasted outcomes to actual events to help the commander determine force effectiveness and measure progress towards achieving objectives.

These definitions highlight the close relationship between the two. The intelligence community considers itself in a supporting role for the assessment of operations (FM 2-0, 1-16). It also considers assessment as a supporting process for sound intelligence analyses or studies (FM 2-0, 1-43).   In addition, they both shape the commander's understanding of current conditions and the effectiveness of his operations relative to his objectives.

How does this play out in the field? In most cases the intelligence and assessment groups work very closely together and the delineation of responsibilities is most likely defined based on the personalities of the leaders of each unit. There may be a tendency for the intelligence unit to focus more heavily on the security dimension, but this is situationally dependent on a combination of factors—immediate need, capability, phase of the campaign, and commander's intent. The assessment team may rely more heavily on the intelligence team for support than vice versa. But in some cases the assessment team's products heavily influence the direction of collection and analysis of the intelligence team.

The best way to manage this division of responsibilities is to ensure that the leaders of both the intelligence and assessment teams have a shared understanding about their individual roles and capabilities, as well as their mutual support requirements. A key element of this partnership is an open sharing agreement that promotes the free flow of information and strong situational awareness of each other's major activities. Without such a partnership there is a very high risk of redundant or conflicting products and recommendations.

Part Two: Assessment Methods

Article Six: Establish a Terms of Reference Document

As the previous article suggested, many terms are used interchangeably or inappropriately by assessment teams. In some cases, the terms are used to convey a very precise meaning. In other cases, the same words are used as synonyms. Unfortunately, the lack of precise meaning can generate confusion in the design of the assessment framework, the analysis of data, or the reporting of insights. Thus, it is worth taking the time to properly discriminate between metrics and indicators, between measures of performance and measures of effectiveness, and between other loosely defined terms.

Defining terms can be difficult because source documents often contradict each other, prior experiences may have generated different perspectives, and theory and application will conflict with each other. One of the best practices for avoiding confusion is to create a Terms of Reference document that best reflects the needs of the current campaign.   In past campaigns, a small team would survey the key documents from participating organizations (NATO, UN, OECD, etc…), reconcile significant differences, and recommend a single definition for each term, gain approval from the senior leadership, and then promulgate the Terms of Reference throughout the assessment community. (The Marthinusssen article in the List of References describes one such effort and provides additional references.) Unfortunately, most teams do not establish a Terms of Reference document until after wasting much time in disconnected debates and publishing guidance that confuses, rather than clarifies objectives. Thus, it is important that assessment teams develop a Terms of Reference document as early as possible.

Article Seven: Build the Assessment Framework Iteratively, Incrementally, and Interactively

Invest a few moments in thinking. It will pay good interest. -Unknown

For an assessment process that will span months or years, assessment teams should build their assessment framework iteratively, incrementally, and interactively. In other words, the

assessment should be developed through repeated cycles (iterative), starting by focusing on a small set of strategic objectives and adding greater detail as needed in every subsequent cycle (incremental), and through collaboration with data collectors, other assessment teams, and decision makers (interactive). This allows the assessment team to refine the focus and scope of the assessment framework based on what was learned at all levels during the development and use of earlier versions.

Overdesigning the assessment framework at the outset can lead to several problems. First, the system may prove to be too complex relative to the available data and capabilities of the assessment team. For a large, multidimensional framework, the associated data collection and management process may be so cumbersome that the assessment cell cannot efficiently and effectively access and analyze the information in time to influence the decision making process. In addition, the identification and ranking of a priori objectives may poorly reflect the real needs of the operational system, with the real needs emerging only after the initial assessment reveals key relationships and trends.

Whenever possible, the process should start with a general assessment of conditions defining strategic objectives, using readily available metrics that help refine our understanding of the roles of these metrics as indicators in key relationships. Each subsequent iteration modifies the metric set and relationships between indicators and conditions to answer key questions arising from prior stages of the assessment process.   This iterative process occurs up and down the chain of command repeatedly as the assessment is refined. See Article Twelve for a detailed practical example.

One potential problem with this approach is that changes in the framework threaten the consistency and continuity of assessment products. Teams minimize this risk by retaining a small core set of metrics throughout the development process for the specific purpose of tracking key conditions and relationships over time. If these core metrics were selected based on the original criteria stated above—readily available, generally applicable—then they can be retained at low cost and will capture the broad trends. As a result, the framework preserves much of the desired continuity.

Article Eight: Discriminate Between Indicators and Metrics

Most people use the term indicator and metric interchangeably, without generating confusion or negative consequences. Even the NATO Handbook, which separately defines "indicator" and "metric," at one point (NATO Handbook, p. 90), later uses the two interchangeably throughout the handbook. Based on these observations and personal experiences of many analysts it appears that we should generally not worry about discriminating between the two.

However, when developing the initial assessment framework, discriminating between indicators and metrics helps focus the selection process, resulting in clearer definitions of measurement requirements and the relationships between observations and desired conditions. During this development process we can define a "metric" as a measurement of the state of one variable or item of interest and use the term "indicator" when we are discussing measurements of data in the context of a relationship—such as an indication of progress toward an objective that is reflected by that particular measurement.   (Kilcullen, 56) Thus, a measurement describing the state of one variable is a metric, but once applied to assess a causal relationship between two

variables it becomes an indicator. For example, the price of tomatoes is a simple and accurate metric measuring the cost of acquiring a tomato. However, it is often used as an indicator of other factors such as the health of local markets, the impact of security risks on transportation costs, or other conditions. Another example related to governance, measures the amount of taxes collected by local government in a district. In isolation this is a <u>metric</u>. But when we use the changes in tax revenues to assess changes in the effectiveness of local government, accompanying increases in loyalty to the local government, and conversely, less influence of local insurgents, then the data is being used as an <u>indicator</u> of improvements in the governance dimension. Note that a metric can be an indicator in many separate relationships, but a metric itself can be defined in very specific and narrow terms, allowing us to more accurately and consistently calculate its value.

This guide primarily uses the term indicator to describe the data used in the assessment process. However, on occasion the more specific term "metric" will be used if there is a need to emphasize the role of a measurement outside any particular causal relationship.

Article Nine: Use Each Class of Indicator Properly

"You know what is wrong with a lot more confidence than you know what is right" — *Nassim Nicholas Taleb, <u>The Black Swan</u>*

Some indicators can be grouped into classes because they share a common set of characteristics that may be beneficial or detrimental to the assessment process. This article describes several of these broad classes including input and outcome metrics, metrics that measure when a condition has NOT been achieved (spoilers), metrics that can indicate positive or negative effects depending upon context (bipolar), and metrics that serve as substitutes for other hard-to-measure indicators (proxies).

Input and Outcome Metrics: Metrics that measure levels of effort and those that measure resulting outcomes both add value to an assessment. The key is to use them properly. Most input measures add limited value as indicators of progress to desired conditions. Examples include money spent on projects and personnel trained. These are typically measures of performance and only tell us the extent to which we have completed actions. In contrast, most outcome metrics assess the effectiveness of completed actions and are very useful in indicating progress towards desired objectives.   Outcome metrics that capture the effectiveness of the aforementioned activities include restoration of potable water to a village and reduced complaints of abuses regarding local, formal security forces, respectively. (NATO Handbook, 1.3.2) While the assessment effort should emphasize measures of effectiveness to determine progress, it should also include measures of performance because the combined analysis helps reveal which activities influence progress the most—reflecting the related measure of efficiency.

Spoilers: Sometimes it is difficult to tell when you have achieved a specific condition required for a transition to occur.   There might be a situation where several indicators suggest that the condition has been achieved, but nothing that is conclusive relative to the risk involved with falsely assuming the condition has been achieved. One method for improving confidence in an assessment that the campaign has attained desired conditions is to examine a class of indicators sometimes referred to as "spoiler" indicators. As the name suggests, these indicators serve as show-stoppers if they reach certain levels. Their purpose is to clearly illustrate that a particular condition does not exist. In a sense, one can think of them as confirming the absence

of a necessary sub-element of a particular condition. In a more statistical sense, they are a test for a false positive finding of achieving a particular condition.

To develop spoilers first clearly define the desired condition and then identify the indicators that are inconsistent with this condition. A good approach to identifying spoilers is to consider the arguments of your critics. Critics usually offer many reasons why they believe the campaign efforts are failing. When properly screened and tested critics' arguments may provide indicators and even evidence of such failure.

A spoiler metric for security would be a lack of trust in the local police and army units. For rule of law, a spoiler might be evidence of significant intimidation or kidnapping of judges. In a broader sense, a spoiler metric might indicate that a province is being used as a safe haven for al-Qaida. Any of these circumstances reflects the lost legitimacy of the provincial or national government in the eyes of the population in terms of the government's ability to provide security or rule of law. Attempting to transition the government lead to the host nation under such conditions will likely result in effectively transferring control of that province to the insurgents instead of to the host nation. Using spoiler metrics such as these should guard against falsely assuming the conditions are supportive of transition.

Bipolar Indicators: Some indicators cannot be interpreted accurately in isolation from other variables. In one environment an increase in the level of the indicator suggests progress towards desired conditions, while in another environment an increase in the level of the same indicator suggests regression from the desired condition.

A simple example of a bipolar indicator is the price of tomatoes in multiple districts. In practice, this metric is used to assess several different conditions, both economic and security-related. A decrease in the price of tomatoes could be due to increased access of local retailers to established supply markets (a positive change) or decreased access of local wholesalers to regional retail markets due to deteriorating travel conditions (a negative change). Whether the price decrease is viewed positively or negatively depends, at least, on whether it is a tomato-supplying or tomato-demanding location, but even this one additional consideration is not sufficient.

In a more complex example, the number of security tips from locals can be bipolar if used in isolation. In Afghanistan, for example, if the local population has low confidence in the local security forces (ISAF, ANA, or ANP) and fears retaliation by local insurgents for cooperation with local security forces, then they are not likely to report insurgent activity and the number of tips may be low. If used to assess the level of insurgent activity in the area, the number of tips reported is misleading. If used to assess confidence in local security forces, the same indicator is accurate. If the number of tips goes up, it may be due to increased confidence in the local security forces or due to a pure increase in insurgent activity, while the number of informants remains constant. In each case, the proper interpretation depends on linking the metric to the right relationship in order for it to be a valid indicator.

A third example is the cost of hiring a local national to plant an IED or fire an RPG. This price can go up or down depending on many things—the number of potential hires (reflecting more or less support for insurgents, the strength of the local employment market, or risk of detection); strength of the local insurgent force (shortage of their own forces, availability of cash, effectiveness of this tactic relative to other tactics), etc. In any case, whether an increase in the

cash payment to plant an IED can be assessed as a positive or negative development depends upon what has changed in many other conditions.

Unraveling the true meaning of a change in a bipolar indicator requires the support of additional indicators and careful disaggregation of the data. If the assessment network is robust, the analyst can essentially query the data, rather than just accept it. To do this, an assessment team can focus on a particular district experiencing the greatest change in the bipolar indicator and by talking to forces in the field build a more comprehensive picture of what is actually happening.

For examples of other bipolar metrics see Kilcullen, pp. 57-58.

Direct and Proxy Indicators: Some relationships of interest can be measured directly. If we want to measure how secure a local official feels in his district, we can observe whether or not he moves about freely without a security detail and whether he lives full-time within his home district. But other critical indicators are just too difficult to collect directly, such as whether this same individual is trusted by his constituents. To make the latter assessment, we need to rely on proxy indicators that substitute for the desired indicator, such as survey data or the number of appeals to other area leaders for help. Few things can ever substitute perfectly for another and this is just as true for indicators. When using substitute indicators, assessment teams need to be clear that the indicator is merely a proxy.

The price of tomatoes across districts discussed above is an example of a substitute indicator. Because it is readily collected across large geographic areas, it can help assess the differences in many conditions across districts such as differences in transportation costs due to security risks or proximity to a healthy economy.

An established best practice is to minimize the use of proxy indicators relative to direct indicators. In addition, proxy variables should be validated whenever possible. For example, if indicators of actual differences in transportation costs or the health of an economy are directly measurable in some geographic areas along with the price of tomatoes, this information can be used to assess the validity of the proxy variable as a substitute for the direct indicator. If this comparison cannot be done regionally, consider trying to validate the proxy variable through some form of cross-sectional comparison, even if it means relying on data collected outside the region.

One of the dangers of using proxy indicators is that teams often continue to use them long after direct information becomes available. To avoid this you should clearly identify proxy variables and reconsider their continued use on a recurring basis.

Article Ten: Beware of Manipulated Metrics

One concern that applies to almost all metrics is that they can be manipulated by a group or individual participating in the activity that is being assessed. Borrowing a term from the field of government economic regulations, we would call these metrics "captured" because they reflect the signals the subjects of oversight want to send rather than reflecting the reality of the current condition. Metrics that are used to promote or demote, or directly redistribute resources and money are at a particularly high risk of being manipulated. Captured metrics can provide misleading information on the effectiveness of governance or local forces or appear to negate

assumptions regarding the relationships between COIN activities and their effects on key conditions.

Some examples from the past include inflated reports of the operational readiness of host nation forces, inaccurate accounting of provincial budget obligation and execution rates, reduced reporting of civilian casualties resulting from local security force abuses, exaggerated reporting of enemy casualties, and "ghost employees" on employment payrolls to increase the amount of development funds distributed through local leaders.

Because COIN is a dynamic, multi-stage learning contest, it is also an information contest. Insurgents have a strong incentive to manipulate information if they know this will mislead, disrupt, or redirect your efforts to their advantage. Disloyal, corrupt, or intimidated officials can also be a source of distorted information in the metrics flow.

One way to recognize a captured metric is to compare rates of change for related variables to see if a causal or complementary relationship appears to be broken. Captured metrics will also be more prevalent where known corruption exists, so we need to actively screen data from these sources.

Some of these problems can be readily overcome by validating potentially captured metrics with complementary metrics—metrics that move in generally the same direction and magnitude as the target metric. Operational readiness reports can be crosschecked with a partnering unit's evaluations of field performance. Provincial budget reports can be validated against program-level execution or production reports. The solution for other problem metrics is to collect the data for assessment purposes at a level as close to the desired final effect/targeted group as possible to minimize misrepresentation of progress or conditions. Avoid culling the data from reporting documents whose real purpose is to evaluate leadership effectiveness or budget competition. Data in these types of reports tend to be less accurate due to the greater rewards obtained by manipulating the data.

More strategically, we need to create the right incentives for providing accurate data. These include tolerance by supervisors for negative news, and relying on a stronger burden of proof for metrics that are used to distribute resources, rewards, and promotions.

Article Eleven: Develop a Manageable Set of Metrics

Not everything that counts can be counted and not everything that can be counted counts.
- Albert Einstein

There are hundreds of metrics available for consideration to support our assessment of progress towards desired conditions.   But data collection can be risky and costly in terms of unreasonable demands on limited resources and time. In addition, since there is a risk that operational teams prioritize their management efforts towards what their supervisors measure, if we do not clearly establish management priorities through a separate command mechanism, a poorly designed metric set may confer unjustified priorities on some activities. For these and other reasons, we need to control how many metrics we use in the assessment process through a form of cost-benefit analysis. (FM 5-0, Appendix H, H-20)

One of the first tests for a potential metric is whether you can consistently **measure** the metric over time. The measurement technique can be quantitative or qualitative, but it must be feasible and consistent.   If it passes this first test then you must consider the costs and risks of

maintaining that consistency in terms of resources required, time spent, and lives and material placed at risk.

The handbooks often address this second issue by asking if the metric is "**collectable**." We collect some metrics specifically to feed the assessment process, while others are gleaned from routine reports designed to support daily operations. The former group must justify the additional resources, time, and risks required for collection solely on the value these metrics add to the assessment process. The latter group can be considered off-the-shelf metrics, incurring much lower collection-specific costs. Clearly, relying more on off-the-shelf metrics lowers the overall collection cost for the assessment process. However, just because a metric is available does not mean it should be part of the core metric set. The total number of metrics used still needs to be minimized since at some point too many different metrics will dilute the analytical effort.

Assessment handbooks establish a third criteria for good metrics--**relevancy**. Some measures of metrics relevancy are obvious. Metrics should announce the exceptional occurrence, and if at all possible, serve as leading indicators for it. Definitions of "exceptional" vary, but a good place to start is to determine what gives the leadership nightmares-what turn of events do they worry about the most due to its impact on the success of the campaign? Many leaders keep explicit lists of these concerns and discuss them with their deputies and staff. In the past, assessment chiefs have directly asked the leadership what worries them the most. Once these concerns are established, a team can design metrics to monitor the trends that culminate in such events and signal their occurrence if they happen.

No matter how good your metrics are relative to the previous three criteria, they must also be available in a **timely** manner. If the metric is not available and or cannot be analyzed in time to meet the decision maker's schedule, then the metric cannot support effective action. To put this issue in the context of Col John Boyd's work, reliance on such a metric places you outside the enemy's OODA loop (Observe-Orient-Decide-Act) and you have lost the initiative (Hammond). If data is reported quarterly by the responsible organization, but key conditions are sensitive to weekly events, then metrics based on this data source may not be useful for tracking important trends and events.

Going beyond the handbooks, there are other factors to consider when constructing a set of metrics. Metrics should **complement** each other in a way that raises the analyst's confidence in his assessment of conditions. Accurate evidence of events often lags the actual occurrence or is greatly exaggerated in initial reports. Measures of civilian casualties suffer from this problem. At the time of a large IED event, exaggerated casualty reports from the street or first responders reach the media and public immediately. There is usually a lag of a day or two before the real casualty figure emerges from hospital or morgue reports. Building sets of leading, lagging, and reliable indicators from complementary metrics improves the accuracy of the data used to measure these types of events.

Another way of limiting the size of the metric set is to **build a diagnostic hierarchy** with your metric set to determine which metrics are collected and reported all of the time and which are collected on demand. Consider a medical analogy. In the absence of a known problem, doctors usually only look at blood pressure, temperature, and weight. Only if a symptom is off the trend line do they look at other measures. In a counterinsurgency assessment we may have more than three key metrics. But all we need are enough metrics to act as a signal if something

happens that is very different from the past. We can then explore the underpinnings of that signal to assess if it is positive or negative. If we think of some metrics as common indicators of instability and some as definitive indicators of a specific source of stability, we now have the start of a diagnostic hierarchy. This system works particularly well when the definitive indicators are a natural part of the operations management process, allowing analysts to look back in time for metrics that were generated and preserved, but not processed or reported. Political and economic activities generally meet these criteria.

For example, two broad economic indicators are employment rates and price changes. Employment may be measured by jobless claims, polling data, business surveys and other means. Price changes can be measured by market surveys, polling data, or shipping records. Any sharp changes in any of these metrics signals an underlying change in the economy which could be related to economic, political, security, or diplomatic issues (see the section on bipolar metrics). Rather than mandating collection of data across all four dimensions in anticipation of need, analysts can collect a small key set of metrics that serve as broad indicators and only target their more widespread collection and investigation on the most likely source in response to the need. Whatever underlying factor has changed significantly should be readily discernable if it was significant enough to affect the broader economic trend.

Article Twelve: Retain Balance in Both Metrics and Method

Several interrelated debates persist about appropriate metrics and methods that should be resolved jointly. One debate argues the merits of the narrative over summary graphics built from aggregated data, with an internal sub-debate over the value of qualitative versus quantitative data. Another debate exists about the strengths of conducting assessments at the lowest level in order to preserve context which is lost as data is aggregated and analysis watered down at higher levels of assessment. Yet a third debate bemoans the loss of the commander's judgment as subjective feedback and frontline views are overshadowed by quantitative reports and color-coded charts in higher level assessment products. Proponents of each argument have valid concerns that need to be addressed. Making the most of an assessment process requires recognizing that the choices are not always mutually exclusive. Strategic assessments can capture the best of both approaches to each of these debates without sacrificing too many of the most important attributes. However, developing such a balanced approach requires a creative, energetic, and proactive assessments team.

Regarding the first debate, the narrative and summary graphics approaches both have inherent advantages. Narrative preserves context. It conveys insights more readily if it represents an insightful synthesis of what the analysis has revealed. But, an assessment can't plot a series of narratives to show progress like data-based graphs, nor can it aggregate narratives across regions without resorting to some color-coding scheme which itself requires quantitative data. The real question is why can't we provide both? The typical answer is that the customer wants a summary—a neat package with one bullet sound bite. In this case, the problem lies with the customer—he is accepting an inferior product in order to minimize the time it takes him to digest the information. The job of the analyst is to win back time from another activity and increase the value of the assessment so the reviewer will readily devote the necessary time to consider both the narrative and the key quantitative data. It helps if the narrative is so strong that the data is relegated to the role of an available reference, rather than the principal information. If you successfully pull this off once by demonstrating that you have both the insight in narrative

form and the supporting argument in quantitative form, the reviewer will ask for fewer detailed data reports in the future knowing that the data exists and has already been scrutinized by the assessment team and is available to support the narrative as needed.

The second debate is a related argument over preserving context. Note that if the first debate is resolved in favor of a hybrid narrative-quantitative assessment process, then the issue of where the assessment is conducted is less important. By retaining the context for the data in a narrative the assessment team can recognize the break point beyond which data can no longer be effectively aggregated. A comparison of low-level narratives for different regions will reveal different trends, causal relations, and concerns of these tactical units. As the assessment team reviews these narratives, they capitalize on these multiple perspectives to synthesize a more comprehensive assessment of progress towards strategic objectives. The choice is not whether the assessment is performed at the low-level or strategic level. A good assessment is developed through interaction that flows up and down in a feedback cycle between the levels until a shared understanding emerges.

The last debate arises from fears that the commander's or LOO owner's perspective is lost when computer-based "effects-based assessment models" collect, aggregate and process quantitative data for assessment reports. The loss occurs because the models are developed in a sterile, deterministic environment, characterized by compromises, hidden subjectivity, and a static view of causal relationships between actions and effects. Such models just can't capture the dynamic nature of the COIN environment or the instincts and insights of those on the front-line who live and breathe the daily flow of events at the tactical level. Much of this is true. But there are few obstacles to preserving the "front line perspective" and using it to assess and communicate alternative interpretations of events or even to reshape the assessment model periodically. The principal requirement here is once again designing an assessment framework that maintains a two-way flow of assessment dialogue as the assessment product spirals upward toward its final state. (FM 5-0, Appendix H, H-26)

Once again, prior experience supports the feasibility of an approach that manages the conflict within all three debates. Prior to the summer of 2007, the Commander's Assessment and Synchronization Board (CASB) in Iraq was mainly characterized by a PowerPoint presentation using both quantitative and qualitative data related to a multitude of issues for five or more hours. There was much breadth and little depth and the process was seen as the **culmination** of the quarterly assessment. Beginning in the fall of 2007, the CASB was redesigned to focus on the major issues requiring synchronization across LOOs. More importantly, the CASB became a **process for assessment**, rather than a culminating event. The massive volumes of data and charts from the briefings were relegated to an Appendix in the read-ahead package. It is revealing that not many principals attending each subsequent CASB read this Appendix. Two other shorter documents were added to the read-ahead. The first was a rudimentary, qualitative stop-light chart that summarized progress toward 10-12 objectives for each of four lines of operation from the Joint Campaign Plan. For each objective, there was a color signifying the need for attention of the principals due to a gap in progress, a two to three line comment explaining the gap, and a reference to the supporting data in the appendix. The most important document was a 3-5 page narrative for each line of operation that discussed progress towards the three most critical objectives under that line of operation. This focused narrative was read and discussed by most attendees. At the CASB meeting itself, each line of operation owner briefed

his or her single-most important issue of those included in the 3-5 page narrative and addressed question on any of the issues in their area of responsibility.

In this form, the CASB was a more effective means for communicating the prioritized concerns of the line of operation owners and for focusing the next stage of assessment. During the CASB discussion, senior military and civilian leaders provided feedback on the line of operations owner's challenges and personal assessment. The assessment team recorded the discussion and decisions of the senior leadership and identified areas for further investigation to validate concerns with empirical evidence or explore potential multidimensional linkages between issues. Over the next two weeks the assessment team worked closely with their line of operations counterparts to (1) summarize the findings of the CASB, (2) augment those findings with additional information to either support or question the findings, and (3) identify issues that needed to be assessed prior to the next CASB in order to synchronize campaign plan activities. The findings and recommendations in the CASB summary was reviewed by the Ambassador and Commanding General and led to shifts in activity levels by line of operations owners, changes in data collection practices, reinterpretation of trends, revised assessments, and the launch of special studies to explore newly discovered issues.

Reviewing the record of the CASB in Iraq, we can see that it incorporated both narrative and summary graphic data, it preserved the context and objective data from lower level assessments, and it ensured that the commanders'/directors' judgment was communicated, reviewed by his or her peers, and assessed against the known conditions on the ground by the best data available. Because it achieved all of these objectives, the CASB assessment carried sufficient weight to drive decisions to reallocate effort and resources within and across lines of operations. Supporting the CASB process was hard, time-consuming, and complicated. But, the returns justified the effort—the enhanced credibility of the information presented led directly to resource decisions by the senior leadership. Many participants chafed extensively prior to the first CASB of this new format because it was a demanding and intrusive process. But, resistance dropped significantly as more and more participants saw for themselves the return for their efforts.

Article Thirteen: Deploy Field Assessment Teams

An informative operational assessment requires both analysis—the deconstructive review of key elements—and synthesis—a creative and instructive integration of related insights to answer key questions regarding the progress of the campaign or effectiveness of critical activities. In conditions of perfect information and seamless connectivity between information systems, a good assessment team can perform these key functions from a centralized position near the top of the operational chain of command. However, in a COIN environment information is not perfect, and information systems are poorly linked, so penetrating analysis and revealing synthesis by isolated staff agencies is rarely possible. Unfortunately, for a variety of reasons, staffs are asked to perform real analysis and synthesis without venturing into the field resulting in detailed, but uninformative descriptions of the current conditions.

A good operational assessment should provide **actionable** information to the audience. To meet this standard, the insights must be relevant and credible. For critical issues, the only way to achieve this standard is to be intimately involved in the data collection and analysis steps.

Thus, the assessment team needs to periodically get out to the frontline units and engage directly with the operational units to investigate the situation on the ground. To understand why this is so important we need to take a closer look at how we perform assessments.

First, it is important to understand how commanders make their personal assessments in their areas of operations. They continually visit field units for briefings and battlefield circulations. The information that has been pushed to them can be explored more thoroughly by personal observation of conditions and two-way personal communication. By immersing themselves in the action they can see how events evolved in response to their direction and develop a natural feel for what is happening. A commander or director who doesn't visit the field quickly loses touch with what is going on. So it is unrealistic to expect an assessment team to remain in touch with what is really happening if it never goes into the field. The media is in the field, the think-tank visitors go to the field, and the commanders go to the field. One cannot expect an assessment team to build informative assessments unless they also go to the field.

Second, some might argue that the assessment team's job is to synthesize reports from others in the field and need not have a field presence themselves. This argument is based on a flawed conception of the objective of operational assessment and how the critical relationship between analysis and synthesis supports this objective. Analysis leads to understanding two things—the nature of the individual parts and the interdependence between the parts. Synthesis begins once the analyst achieves this understanding. It melds the analytical insights in novel combinations to create new concepts, solutions, or realities. (Dettmer, Part 3) The successful art of synthesis requires participation in the act of analysis. One cannot create novel combinations that **inform** the audience in a way that increases its influence over their environment without a solid understanding of each dimension of the environment. This solid understanding can only come from direct interaction and investigation of events where they happened—in the field.

Lastly, assessment should not be treated as a historical analysis. COIN takes place in real time, with real participants, with whom assessors can, and should interact. Assessors do not have to wait for the final stage of an operation to assess its progress. As events unfold, assessors should look for leading indicators to assess if the operation is on course with the strategy. Assessors are not limiting to theorizing about theorize cause and effect or running statistical tests on key variables. Since most variables of interest relate to human behavior, it is not difficult to directly ask the people involved about unfolding events. There is no need to prove hypotheses with statistical sampling over large data sets, we can query our "test variables" directly to understand their behavior, and **know** what happened, even with small samples. The only way to develop this investigative relationship with our subject matter is to interact directly with those in the field.

In practice, this approach is similar to what MG Flynn proposed to improve intelligence collection. We need to create teams "empowered to move between field elements, much like journalists, to visit collectors of information at the grassroots level and carry that information back with them to the regional command level." (Flynn et. al.)

Flynn's journalist analogy is instructive in several ways. First, journalists go to the field to engage directly. They prefer to witness events and use direct interviews, not second-hand accounts. Second, journalists recognize that it is impossible to cover every issue, district, or operation. They prioritize and cover only the critical ones. Assessment teams can do the same, dedicating one or two experienced analysts to respond in near real-time to conduct a field

assessment for critical events. Examples of critical events would when a newly operational host nation military unit participates in a major operation, when new sources of information become available by opening of a new Combat Outpost, when an unusual pattern emerges in polling data, or when there is an important meeting of PRTs and local leaders to discuss provincial development plans. Each visit should seek to establish a keen understanding of the current conditions through key metrics and qualitative records and report the findings in narrative form.

This "expeditionary" mindset improves the assessment process in many ways. Field assessment teams can gather detail that is missing from regular field unit reports. Teams will gain a better understanding from the data collectors regarding how they translate qualitative observations into quantitative data. It allows for "sighting corrections" through discussions with collectors so both collectors and assessment teams learn to interpret observations the same way. Field testing the assessor's interpretation of data also helps refine collection targets to better match the analysis of issues and lines of critical interest. And again, this direct interaction allows field units answer key questions directly, eliminating the need for the analyst to deduce an answer from limited data. Finally, field visits restore the context that may have gradually divorced itself from the data as it was sent to higher levels.

While field assessment clearly adds considerable value to the final product, we must minimize the burden placed on host units by these field assessment visits. Given the current lack of understanding of field data by assessment teams, the first large gains in understanding should come at little cost to host units. First, analysts should travel to units when those units are already briefing visiting groups (academics, congressional delegations, think tanks). In 2007, assessment staff analysts in Iraq shadowed the General Accounting Office (GAO) team for eight days as they visited division staffs and some front-line units. The analysts returned with a vastly improved understanding of conditions on the ground and the information available at host units to credibly support the operational picture. The team repeated this process many other times with equally impressive results. Alternatively, analysts can learn the schedule of major meetings occurring on various issues and attend as observers. Assessment team members in Iraq attended PRT planning meetings, host nation provincial coordination meetings, commander's conferences, and a host of VTCs. They acted as recorders, rather than interlocutors, reserving their questions for post-meeting discussions. Analysts should also look for opportunities to serve their host's needs. As they develop relationships they learn of issues that are too large for the subordinate units to handle. In Iraq, analysts were occasionally placed in a unit as a liaison or action officer. This let the analyst "swim" in the formal and informal data stream of the host unit. Not only did the analyst help the host unit understand their environment better, but the assessment team learned what type of information was available and how the local commander's used it to interpret local conditions.

These are just a few of the benefits from occasionally putting teams in the field. These visits are costly in manpower terms, but they would always pay off handsomely in both short- and long- term improvement to assessment products.

Article Fourteen: Bound Estimates with Eclectic Marginal Analysis

Assessments mostly focus on using estimated metrics to test relationships between actions and effects, rather than measuring key relationships under controlled conditions. This is true of most aspects of any social science—and many aspects of fighting a counterinsurgency

have been described as the conduct of "armed social science". Assessments are supposed to measure the state of certain conditions, but given the nature of the work in a counterinsurgency—i.e., trying to examine social relationships—we have to rely more on art than science in the measurement process.   In addition, because social relationships operate through and across networks with multiple dimensions, the analyst should expect that a change in one condition would trigger observable responses throughout the system. Recognizing these aspects of assessing counterinsurgencies, one can use marginal and eclectic methods to approach the problem from a different perspective.

In general, when assessing progress towards certain objectives we should ask ourselves, why we expect anything to have changed from what it was in the prior period, unless something significant (and observable) has happened. The default assumption should be that nothing has changed. This is referred to as the marginal approach. We should also ask ourselves why we believe anything has changed from what it was in the prior period unless there are multiple indicators that it has. Then, by examining the relative changes in this set of indicators, we gain an idea of the magnitude of the change—the eclectic bounding approach.

There are other reasons for using marginal and eclectic techniques. Collecting data on some indicators is extremely difficult, particularly, if one tries the direct approach and attempts to find a metric that is individually conclusive. The problem can arise for many reasons. The metric might have been designed perfectly, but it is uncollectable due to lack of access or resources required, or it may be measurable only a long time after that event occurs. In other cases, no single metric is conclusive, so a set of metrics is needed in order to properly interpret the condition or validate indications of a critical shift in the environment.

Measuring progress indirectly using the idea of marginal and eclectic analysis can also sometimes yield promising results. When measuring indirectly, the focus is on complementary indicators that help us bound the problem, not on the key metric itself. Marginal analysis does not estimate the value of the metric: it estimates the amount by which the metric has changed since its last reliably-known value. In marginal analysis one assumes that nothing has changed unless a significant force or event influences the target variable. The key to a successful marginal analysis is identifying these influential relationships. The examples below serve to illustrate these concepts.

Commanders always want a rough order of magnitude estimate of the size of the opposing force. Given the more global nature of the current insurgencies, there is particular interest in the number of foreign fighters among the insurgents. Just to make a single estimate of this number for a particular period of time takes an enormous amount of intelligence resources. The magnitude of this effort precludes frequent updates through the same process. Rather than repeating the process for monthly updates, the assessment team can monitor other more readily collectable metrics to develop a current estimate. Monthly figures on foreign fighters killed or captured, HUMINT reports on disenchantment and defection of recruits, and estimates of the capacity of the foreign fighter pipeline are several data points that capture the inflow and outflow from the stock of foreign fighters.   Putting them all together generates an estimate of the current stock of foreign fighters without having to repeat the original process. This type of rudimentary, marginal analysis helped keep estimates of foreign fighter forces in Iraq in 2008 more up-to-date. It can also be used to keep other similarly complex indicators on target in between rigorous, bottom-up intelligence estimates.

Another indirect method for completing an assessment is to bound the indicator of interest by an eclectic set of complementary indicators. This bounding provides increased confidence in the accuracy of the assessment. For example, when trying to demonstrate the effectiveness of the Sons of Iraq program (also known as Concerned Local Citizens) the assessment team used this method to show the real value of the program relative to its costs and risks. The objective was to show that hiring former insurgents under the direction of a local tribal leader could improve local security. There was no way to statistically prove this case; however by tracking a set of seven related indicators (irrefutable reductions in civilian, coalition, and Iraqi Security Force deaths, armored vehicle battle loss costs, IEDs exploded and increases in IEDS Found/Cleared and Caches Found/Cleared) with the growth of the Sons of Iraq program the team built a persuasive case that this $5M/month program was responsible for saving hundreds of lives and over $10M in equipment costs each month. The argument was further supported by three accompanying anecdotes to demonstrate the causal link between the recorded results and the Sons of Iraq program. The summary slide for this assessment is provided on the next page

Applying these techniques requires creativity. Assessment teams should seek novel ways of combining metrics to paint what looks like a collection of dots to the statistician, but reveals itself like an impressionist painting of current conditions to the decision maker. The advantage of this approach is that it forces us to recognize that we are dealing with a dynamic social system, not a controlled experiment. The resulting assessments avoid overstating accuracy and change, while providing a clearer picture of progress.

.
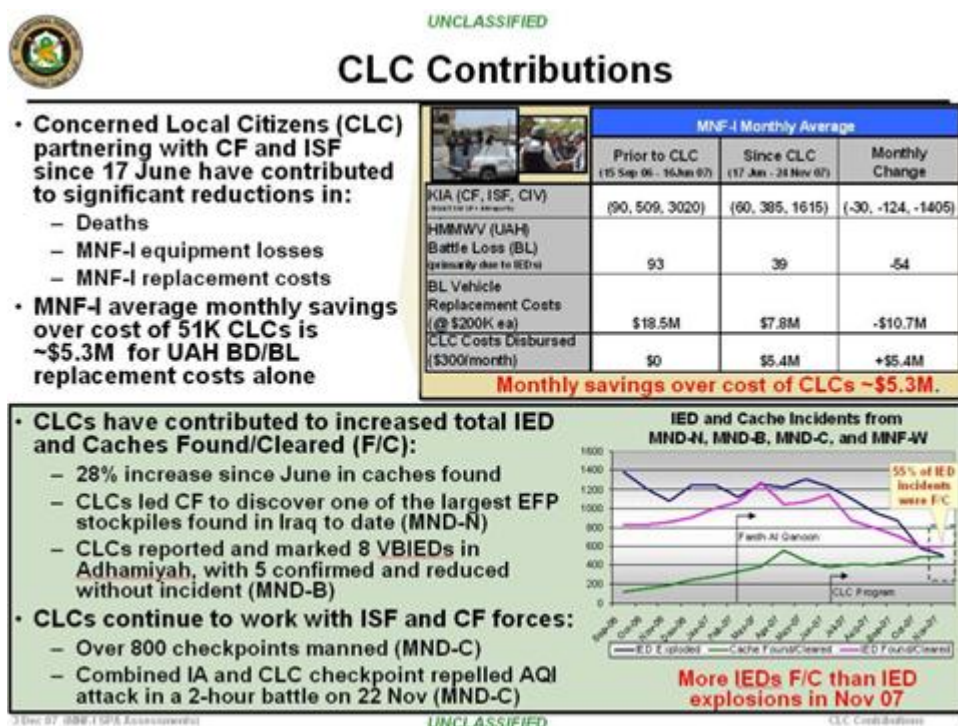
Article Fifteen: Anchor Subjectivity

The eye sees only what the mind is prepared to comprehend. - Henri Bergson

Critics of qualitative indicators often point to the significant degree of subjectivity inherent in this type of indicator. This criticism is generally warranted. However, rarely do we have the option to substitute a precise, consistent quantitative metric for the qualitative indicator. The underlying data is just not available. Thus, analysts' efforts need to focus on minimizing and anchoring the subjectivity in the qualitative indicator. They should also consider that the degree of accuracy required in the indicator is tied to the way it will be used. Specifically, most qualitative indicators serve as warning signals, much like dashboard lights in a car. Their primary function is to indicate whether we are in or out of the normal operating parameters for the system. They can also be used to confirm quantitative indicators. They don't have to tell us precisely where the campaign stands, just whether it is far from the target or from the previous position. Their purpose is to tell the analyst when to check on something because it looks abnormal. By considering both the bounds for normal readings and the sensitivity of the indicator he can design qualitative indicators to minimize and anchor subjectivity.

When considering the type of indicator to use, first define the normal operating parameters for the benchmark metric associated with the condition of interest. The benchmark may be the current level in a specific geographic area, acceptable levels already achieved in other geographic areas, or it may relate to internationally-defined standards such as those used by the World Bank or other UN agencies. Establishing normal parameters for this benchmark include the moment value (the current value) and bounds to define what are considered significant fluctuations around the moment value. Next, consider whether a quantitative measure is available for the benchmark and whether it is accurate within the bounds of what are considered normal operating parameters. Finally, use the defined bounds for significant fluctuation (similar to standard deviations) to define thresholds or milestones on the path from current conditions to the desired or acceptable level. If all these criteria are met, then there is a good case for using a quantitative metric. If some criteria are not met, this is simply qualitative data and you just have to make the best of the available information. (NATO Handbook, 6.1.2)

In the absence of the criteria for good quantitative metrics outlined above, two options are available. If the problem lies with the accuracy of the metric within its defined bounds, one might be able to build an eclectic set of quantitative metrics that helps "triangulate" your estimate of the benchmark's true value (see the previous article on Eclectic Analysis for an example). If the problem lies elsewhere, one needs to focus on minimizing and anchoring the subjectivity of a qualitative metric.

Minimizing and anchoring subjectivity works in a fashion similar to the process for a quantitative metric described above. First, carefully describe the proposed relationship between the qualitative indicators and the conditions of interest in writing. Be sure to consider each relationship's strengths and weaknesses. A best practice is to share this description widely and to refine it based on feedback. This process also helps define the benchmark measures and develops the bounds of significant fluctuation and thresholds for the desired levels. The resulting product is a type of rubric that can be used by evaluators as they provide their field assessments, helping them remain consistent with the original intent of the metrics. Such a rubric was used in Vietnam to improve the Hamlet Evaluation System (Gayvert, p. 6).

If the qualitative metrics will be aggregated or weighted, gather information on the desired weights through this process as well. The end product from this stage is a **written,** well-defined relationship between the sets of qualitative indicators and their target conditions, as well as a range of estimates on the relative importance of each individual indicator. This process of

development is a hybrid between the Delphi method and the first stage of development of the DynaRank model—a decision support model developed by RAND and used in the Quadrennial Defense Review. The Strategic Assessments Cell in Iraq used this technique in 2008 to reduce subjectivity in their assessments of political progress in Iraq. (FM5-0, Appendix H, H-27)

Any decision to proceed with full implementation of this or similar multiple- attribute-decision models should not be taken lightly. There are many well-known problems with large-scale aggregation of qualitative indicators. Most of these problems occur at the next stages of model development—burdensome collection costs, improper aggregation, diminishing context, loss of insight, and a black box infatuation with the numbers produced—to name a few.   These types of models do not anchor subjectivity, they hide it.

However, the value of this first stage of model development, in which relationships are defined and vetted, is significantly underappreciated and poorly leveraged. In this first stage, a good team can capture the diverse perspectives of leaders and subject these perspectives to peer review. Not only does this process clarify what is important to the assessment, but like the Delphi method it also constructively shapes the leaders' perspectives to narrow their own range of subjectivity. More importantly the remaining subjectivity becomes anchored in the assessment framework in a **transparent** manner so staffs throughout the operational chain are more aware of the sources of subjectivity and the interdependence of the many components of the campaign plan For example, if the consensus is that public support for the local police is twice as important to achieving local stability as the ability to provide reliable electricity, then the assessment team has some guidelines through which they can interpret results that match the leadership's perspective. They should also test the validity of such a relationship by watching the relationship between these two indicators and other indicators of stability, providing feedback to the leadership on the accuracy of their stated perspective.

Another key point to keep in mind in the debate over subjectivity is that just because a number is expressed as a data point does not mean it's not a subjective qualitative measure. Many times field observations are translated to quantitative ratings or thresholds which are set subjectively. For example, the number of villages with adequate access to potable water is reported as a number, but collected subjectively since "adequate" is a composite judgment defined by cleanliness, distance, security risk, and reliability to name a few. Knowing how the question will be used helps anchor the subjectivity. If the relationship is between adequate water and support for local governance, "adequate" is defined in terms of whether this is a local point of contention between the population and its leaders. However, if the relationship is between adequate water and the ability to sustain agriculture then "adequate" takes on a different meaning. If the data collectors know how the information is used then they can refine their thresholds such that the information is meaningful as it flows up. Ensuring that the collectors and the users share why and how they view the data is critical to anchoring subjectivity and this process is facilitated by more direct interaction and transparency.

All of the above describes best practices for controlling subjectivity as analysts develop the assessment framework. But much of this effort is wasted if it is not documented and promulgated to those who will conduct the assessment. Therefore, it is essential to record the enhanced understanding of the relationships between metrics, indicators, and conditions arising from the development process and preserve this information to guide those who will use the metrics and indicators to develop the actual assessments.   This is not difficult to do, but it just requires a dedicated effort. During a 2010 workshop to develop metrics and indicators to

measure improved governance in Afghan provinces and districts, a multinational team of subject matter experts debated key issues, shared their extensive experiences, and gained a more accurate and practical understanding of the challenges related to measuring effective governance. To pass this knowledge on to the eventual users of their products, the team produced a short narrative describing the relationships between the recommended metrics, indicators, and conditions; the potential strengths and weaknesses of key indicators, and the role of any spoiler metrics in the assessment process.

Article Sixteen: Share Data

Every coalition effort faces information sharing challenges due to the diverse demands of national reporting chains, multiple levels of security, and technological constraints.   It is unlikely that this problem can be resolved to everyone's satisfaction, so the real question is finding ways to live with the problem. Fortunately, analysts do not need to share everything with everyone all of the time. Their objective should be to have information that is accessible in an accurate and timely manner when it is needed.

There are two key purposes for sharing information. The first, and most challenging, is to allow for centralized aggregation of compatible data to support strategic assessments across the theater. The second is to allow for targeted assessments to support critical strategic inquiries. The first is a complex task and will most likely suffer from the widely-recognized problems associated with centralized aggregation of assessments. Experience suggests that there is a better approach to strategic assessments that does not require one consolidated database (see Article Twelve). The second objective, strategic inquiry, is achievable and will bear the most fruit.

There are several steps required to share data in support of strategic inquiries. First, it is important to know what information is generally collected.   Second, the parties need to agree on what information they should share more freely. There are several agreements already in place to define such partnerships, but the prevailing trend is towards stove-piping control of data, rather than a presumption that all can share freely. Third, access to this information needs to be streamlined through a network of knowledge managers by developing rules of engagement that minimize the costs of sharing information. While some may read into this the desire for a technological solution, that is not the intent. The intent is to promote a more open collaboration between assessment teams across the theater.

An example of the first task is the Afghanistan Data Cards effort that seeks to create a catalog of data sources that is periodically updated and shared across the assessment network in-theater.   Knowledge managers already know how to do this. The trick is to get them together to consolidate the information in one user-friendly location.

The second task may prove to be very difficult. For a variety of reasons, agencies are protective of their data. They may fear that released data will be misunderstood, intentionally manipulated or misused, or used by others resulting in surprises to the owning agencies. In other cases, they may believe the data is harmful to their own objectives so they want to deny this information to those advocating other views. This is a command issue. Commanders and directors need to set the standards for how much tolerance they have for data sharing within the coalition and they need to be transparent regarding these standards so everyone knows where the lines are drawn. Defining the rules of engagement properly (step 3 below) should give the directors enough confidence to be more open with their data sharing agreements.

The third task, establishing the rules of engagement for data sharing, should rely on three principles. First, requests need to be defined in terms of the question to be answered, not merely as a request for a set of metrics. By specifying the question, the collaborating units know how the data will be used, can offer alternative suggestions or off-shelf analyses already completed on that subject, and can effectively partner with the requesting analyst to ensure the data are interpreted properly. Second, the supporting unit retains the first right of disclosure of key results related to their area of operations. Since they are partners in the assessment process and will always have the most up-to-date information this should be easy to achieve. Keeping the associated chiefs or directors of staff informed is an essential element in preserving this right. However, these supervisors should be strong proponents of sharing and not unduly restrictive of the free sharing and discussion of raw information within the assessment community.   Finally, the final assessment product should be posted on a shared site for future use by the entire assessment community. This facilitates discussion and learning within the community and may help avoid redundant efforts.   For more suggestions on rules of engagement for data sharing see Flynn's article, "Fixing Intel".

Article Seventeen: Include Host Nation Data

Two features of the COIN assessment environment that should be considered when developing the assessment process are the existence of host nation data collection efforts and the ability for assessment teams to interact with this system. Most coalition analysts put little faith in host nation data collection systems. Typically, these systems provide incomplete coverage, devote little effort to validating the data, and can be corrupted by sloppy field craft or political agendas. On the plus side, host nation data sources are often the only system available in some regions and for some topics. In addition, these reports are often developed through more direct contacts with the population. Finally, they reflect what the host nation sees and can provide insights into why the host nation responds to current conditions the way they do. This is very important since in many cases perceptions are often more influential on decisions than reality. As an added motivation for starting to work with host nation data and improving its quality, we need to remember that ultimately we transition ownership and control of all reporting functions to the local government as counterinsurgency efforts mature. If that transition is to be successful the host nation must be able to conduct its own accurate assessments of conditions in the transitioned areas.

For various reasons, access to host nation systems can be problematic. The host nation may be reluctant to share, technological issues may preclude direct links, and data may filter up slowly.   But by directly interacting with the host nation collection team you can overcome some of these obstacles.

In Iraq, during the Basra offensive of 2008, Iraqi forces were in the lead. Coalition reporting was minimal. Media reporting of civilian casualties was sensational.   SIGACTs provided little insight since neither the few coalition forces on-scene nor the numerous Iraqi forces were submitting more than minimal reports. The media estimated over 1,000 civilian casualties in the first few days and SIGACTs reported much less than 100. Pressure was increasing to halt the Iraqi forces offensive due to the high number of civilian casualties. To narrow the range on the estimate of civilian casualties, the MNF-I assessment cell tapped into the Iraqi hospital network in Basra through a US doctor in the Green Zone and his connections with Iraqi doctors. Tracking hospital morgue cases by phone, they were able to narrow the estimate to

more realistic numbers and were even able to discriminate to some degree between civilian and insurgent deaths. These new estimates proved to be much more authoritative and were extremely helpful in dismissing media reports of disproportionate use of force in the vicinity of civilians.

Given the large number of metrics involved in a typical campaign assessment, there is normally not enough time to handle all our data collection needs in such a labor-intensive fashion, but for critical, time-sensitive data, assessment teams should consider using a similar approach. The more assessment teams work with host nation reporting systems, while the international community is in the lead, the better these systems will be by the time the host nation is in the lead. The Afghanistan Data Cards initiative, through engaging Government of Afghanistan representatives, also revealed that coalition forces were simply unaware of many data sources available through the host nation.

Article Eighteen: Develop Metric Thresholds Properly

Thresholds can be used at all levels of the assessment process. Tactical units may establish thresholds that help them report progress on local conditions such as the availability of essential services or levels of violence. Strategic assessment cells may establish thresholds to determine when to transition between different phases of operations or transfer greater control to the host nation.   Thresholds need to be well-designed since minor deviations in threshold values may lead to significantly different assessments of progress. Because thresholds are contextually-specific it is hard to establish one detailed set of rules for their use.   However, we can establish some useful guidelines.

First, like the metrics themselves, thresholds need to be oriented towards the objective conditions within the strategy or reporting mechanism.   For example, a strategic threshold that governs the transition process for security lead might be that violence in the province is low enough that local security forces can independently restore stability despite that threshold level of insurgent attacks. Here, stability could be defined by indications that defections from local forces remain low, that private militias do not emerge, and that trust remains high in local security forces. These threshold levels are relevant because they support the strategic objective—the host nation can successfully provide security at those levels of activity. A tactical threshold for the level of violence in a district may relate to an economic development objective—is violence low enough that PRT, host nation ministries, and aid organizations can proceed with planned development projects? It is important to note that this last threshold is a cross-dimensional threshold—a condition achieved in the security dimension supports a significant change in activity in a separate dimension.

Second, to ensure accuracy and consistency in reporting metrics against thresholds, each threshold needs to be carefully defined and users need to know and apply these definitions. This is particularly important for the tactical thresholds for individual geographic areas because these metrics are often aggregated regionally or may be confused with other subjective metrics. For example, the violence in one area may be reported as positive in the development domain since it is low enough to allow development projects to proceed, but it could be reported as negative in the security domain because it is not low enough to allow the transfer of security responsibility to the host nation. The tactical agency needs to clearly understand the criteria against which they are expected to report, and that linkage between the metric and the different thresholds needs to be retained and understood by users throughout the assessment process.

The first criterion above describes the significance of a threshold in operational terms. As a third criterion, observances of indicators above these thresholds must also be statistically significant. As mentioned in the article on anchoring subjectivity, many metrics fluctuate widely in the short term. This fluctuation affects our ability to develop an accurate measure of the metric and to recognize what are significant deviations from the norm. When setting thresholds, one needs to set them at levels that ensure that a threshold breach is truly a signal of a significant change in the underlying conditions measured by the metric.

There are many sources available to support the development of threshold levels. Since thresholds are heavily dependent upon environmental conditions, the best sources are cross-sectional comparisons within the country or region. For example, to determine tolerable levels of violence to support transition of the security lead for a province, it is best to compare the levels in other provinces where the host nation already has security lead. For developmental thresholds, there are a variety of international organizations that track developmental progress of countries in the same region. These sources are generally not timely enough to provide metrics or short term trend information since most are reported annually. However they may be useful in setting benchmarks for regional quality of life standards.

Finally, while developing, evaluating and proposing alternative threshold levels they need to be tailored to and approved by the unique set of decision makers who will use them. Remember that a key element from our definition of the assessment objective (Article One) is to "provide feedback that influences the decision maker's behavior". Each of the line of operation (LOO) owners will use your assessment products to support decisions regarding different elements of the campaign plan. It is highly unlikely that one set of thresholds will adequately serve all LOO owners equally well.

Article Nineteen: Avoid Substituting Anecdotes for Analysis

One death is a tragedy, a million is a statistic. —Joseph Stalin

Anecdotes are a useful component of assessments when used properly, particularly when used to illustrate a verifiable relationship or reinforce a message. Unfortunately, in some cases they become substitutes for a solid assessment. The best rule to keep in mind when using anecdotes is that they are generally the starting point for analysis, not the closing argument for an assessment. Analysts should test any anecdote before including it in an assessment.

Anecdotes are compelling because they are based on first-hand accounts and are typically accompanied by a colorful narrative. However, in isolation, anecdotes need to be viewed merely as a record of an isolated incident, not evidence of a wide-spread trend. The job of the analyst is to deconstruct the anecdote to understand what key relationships lay behind the narrative, what metrics would confirm the relationship, and where to find evidence of a matching broader historical or geographic trend.

The lack of observable and reliable data drives much of the reliance on anecdotes as evidence. Thus, establishing a solid data collection process to support the assessment framework should help minimize the reliance on anecdotes. In the absence of a robust data base, the analyst can look to confirm the implied trend through some of the techniques discussed above (Proxy Indicators, Eclectic Marginal Analysis, and Field Assessment Teams). The Iraq team used the field assessment approach repeatedly to explore the veracity of the latest anecdotes. By talking with those closest to the story they explored the underlying relationships, searched for

reoccurrences of key events, and gathered direct or proxy indicators to build a compelling case for the trend illustrated by the anecdote.

Once the case to support the anecdote is developed, the analyst should refine the overall assessment framework (metrics, relationships, collection process) so it will routinely monitor and assess the underlying trend using the methods developed for this specific case. This last step is an integral part of the iterative, incremental, and integrated assessment process outlined in Article Seven.

Article Twenty: Use Survey Data Effectively

One of the more divisive debates in the operational community concerns the value of surveys in an assessment process. The common perspectives that "people often vote with their feet" and "actions speak louder than words" suggest that our primary method of assessment should be to directly measure the actions of the people. However, questions of motivation, satisfaction, degrees of trust or fear, as well as intentions regarding future actions are difficult to measure by monitoring actions. The population traditionally expresses the latter information types verbally. Sometimes interviews or broader surveys are the only option for capturing this information.

Another argument in favor of surveys is that analysts should capitalize on the opportunity they have to directly query the objects of their assessment (the people). Analysts can speculate extensively about particular developments and their causes. But it is better to seek valuable information directly from the people themselves.

The major arguments against using survey data in campaign assessments center on issues of intentional bias from respondents due to the potential negative consequences from speaking freely, the accuracy of sampling methods, and the wide range of subjectivity inherent in the questions and responses. No matter how hard we try to minimize these problems they will always exist to some degree. The question to answer is whether analysts can obtain actionable data from surveys at reasonable costs despite these methodological flaws. Given the widespread acceptance of survey data to help understand political campaigns, it seems reasonable that survey data could also contribute to a better understanding of a counterinsurgency campaign.

Experience has shown that we already know how to design surveys that can effectively and efficiently contribute to counterinsurgency assessments. For a detailed look at this subject refer to the Government Accounting Office publication, *Developing and Using Questionnaires*. The more pressing issue covered by this article is how to gain the most accurate insights from survey data once it is collected.

First, users must recognize that in a hostile environment survey respondents may be reluctant to give honest answers. Respondents are typically hesitant to choose sides on an issue like expressing trust in the government when they are talking with strangers—and most survey team members are strangers to the respondents. Before using any survey data, analysts should review the list of questions to see which are likely to have a high bias due to fear of retaliation for an honest answer. But even, when bias is suspected, they should not reject the information completely. The bias is actually a proxy indicator for the people's lack of faith in local security, and that information may be more valuable than the original intent of the question.

Second, one should be careful about statistical significance when reporting numbers, without becoming obsessive about it in your analysis process. Survey data are not measures drawn from a controlled experiment. Instead, the data is drawn from the opinions of people living within a dynamic and sometimes hostile social structure. The survey is not conducted under conditions of stability; the survey is looking for signs of emerging stability. In the analysis phase, analysts should be looking for warning signs, clues, potential trends, and hints that something is changed, to spot the first signs of emerging trends, problems, or success. Minor clues are likely to appear in the survey data on one or two related questions. From these minor clues it is possible to build an investigative strategy that develops more robust, conclusive data through comparing answers to complementary questions, augmenting survey data with quantitative data, or following up with focus groups or field interviews. If the focus is too heavily on precision and statistical tests at this stage you may overlook some key clues. With all that in mind, it is important to avoid overreacting to every minor change in the data trend. As in all other assessment methods, a balanced approach is preferable.

Third, population demography is a critical factor in survey results. Surveys tend to aggregate results and rarely parse them out beyond geographic regions or ethnic groups. But even within these groups there can be important age or social class distinctions that are obscured when the results are aggregated. Younger populations may speak their mind more freely. Professional classes may have more influential local social positions. Aggregation drives every metric towards the mean—but there is rarely an "average citizen" in that aggregated group.

Fourth, assigning the proper role to surveys when you design your assessment strategy is crucial. Surveys are better tools for questioning or rejecting assumptions or theories than they are for confirming them. Surveys are great tools for exploratory work—for example, to widely sweep the environment for promising lines of investigation or broad indications of conditions. But surveys should always be followed-up with more focused investigative practices once the key question has been identified and refined.

Finally, be deliberate in how you report survey results. The focus must be on the insights and trends, not on the numbers, or claims of substantiation of causal relationships solely on the basis of survey data. One should also be conservative in accounting for your margin of error. Textbook statistical rules clearly define how to estimate margins of error, with many polling companies reporting margins of error near +/-3%. But these rules and resulting low error margins are based on assumptions about the survey population or process that may not hold in a COIN environment. One way to compensate for such faulty assumptions is to address the robustness of the findings themselves with statements such as, "this recent trend remains noteworthy even with a +/-10% margin of error."

The bottom line is that survey data remains a valuable component of any assessment framework when used properly. Survey users should understand the most common biases in survey data so they can use the information properly. It is well worth the time spent managing the periodic false positive signals that surveys may send because a survey may be the only way to get an early warning that something has gone seriously amiss with the counterinsurgency strategy.

Parting Comments

Conducting assessments at any level is a challenging task. No guide or handbook can pretend to convey what to do in any great detail because what analysts do in practice depends on what they have to work with in terms of time, resources, and information. In this light, a successful assessment team must act more like a team of craftsmen and less like a team of technicians with set procedures and fixed materials. Analysts possess a diverse and powerful set of tools, but the material upon which we work—the data, population, and environment—changes every day. The analyst's job is use the analytical tools to shape this ever-changing information set to meet the needs of our customer—the decision maker.

Rather than prescribing specific procedures, this guide tries to help analysts think about what they must produce in order to deliver a credible, transparent, and relevant assessment from the available information. By providing a clearer statement of the assessments' purpose, exploring the qualities that enhance the value of assessment products, and examining the key elements of the assessment process, the guide has highlighted practices which enhanced previous COIN assessments. To put all this to effective use, analysts must now create novel ways of using these ideas and methods to improve their own assessment product.

Ultimately an analyst's success will depend upon his or her ability to innovate. Most likely you will never have everything you need, so you should creatively use what is available. Leadership is the art of the miracle, not the mundane, and this is true in leading assessments as well. Keep your purpose in mind. Preserve the integrity of the assessment product, and focus on providing actionable information.

Good Luck!

A Note of Thanks

I would like to pass on my thanks to the many contributors to this product. Principal collaborators on this project include Jeff Appleget, Jim Bexfield, Tom Cioppa, Ben Connable, Lee Ewing, Geoffrey Hartmann, Gary Harless, Gerwin Hennig, Elin Marthinussen, Anton Minkov, and Brad Pippin. We should also thank all those who served with us in the field in Iraq and Afghanistan over the past decade--these are the lessons we learned together. Finally, I wish to personally dedicate this guide to the next generation of operations analysts, particularly Jamie and Will. Take these lessons and build on them.

***Do not confine your children to your own learning for they were born in another time.
- Hebrew Proverb***

List of References and Useful Sources

Connable, Ben, "An Alternative to Effects-Based Campaign Assessment in Afghanistan." PM-3618-MCIA. September 2010. Prepared for the Marine Corps Intelligence Agency

Cordesman, Anthony. "Afghan Metrics: How to Lose a War – and Possibly How to Win One." 15 January 2010, <www.csis.org>.
Department of the Army. FM 2-0 (34-1) *Intelligence*. 17 May 2004

Department of the Army. FM 3-0 *Operations*. 14 Jun 2001

Department of the Army. FM 3-24 *Counterinsurgency Field Manual.* 2007

Department of the Army. FM 5-0 *The Operations Process*. Chapter Six and Appendix H. March 2010

Dettmer, H. William. "Destruction and Creation: Analysis and Synthesis*"*. 27 December 2010. <http://www.goalsys.com/systemsthinking/documents/Part-3-DestructionandCreation.pdf>

Flynn, M.T., M. Pottinger, and P.D. Batchelor. "Fixing Intel: A Blueprint for Making Intelligence Relevant in Afghanistan." Center for a New American Security. January 2010.

Government Accounting Office, Program Evaluation and Methodology Division. Developing *and Using Questionnaires.* < http://archive.gao.gov/t2pbat4/150366.pdf>

International Security Assistance Forces. *Joint Framework for Inteqal*. December 2010

Gayvert, David. "Teaching New Dogs Old Tricks: Can the Hamlet Evaluation System Inform the Search for Metrics in Afghanistan?" Sept 8 2010. smallwarsjournal.com

Ghani, Ashraf and Claire Lockhart. *Fixing Failed States*. Oxford: University Press. 2008.

Hammond, Grant T. *The Mind of War: John Boyd and American Security*. Wash D.C.: Smithsonian Institution Press. 2001

Kilcullen, David. *Counterinsurgency*, New York: Oxford University Press, 2010

Marthinussen, Elin, Bard Eggereide, Frode Rutledal, and Alf Christian Hennum. "Progress Assessment in a Multinational Operation—a Norwegian Perspective." 4th IMA Conference on Analysing Conflict Transformation, University of Oxford. June 2010.

Nagl, John A. *Learning to Eat Soup with a Knife, Counterinsurgency Lessons from Malaya and Vietnam.* Chicago: University of Chicago Press. 2005.
NATO. *Assessment Handbook,* (NATO UNCLASSIFIED). 16 June 2010
O'Hanlan, Michael E. and Hassina Sherjan. *Toughing It Out in Afghanistan*.   Washington D.C: Brookings Press. 2010

Schroden, Jonathan. *Fixing Assessments in Afghanistan* (Unclassified Version). (CNA representative to USCENTCOM). 17 February 2010
Taleb, Nassim Nicolas. *The Black Swan*. New York: Random House. 2010

Thayer*, Thomas C. War without Fronts: The American Experience in Vietnam*. Boulder, Colo.: Westview Press. 1985.

Dave LaRivee is an Assistant Professor of Economics at the United States Air Force Academy. He retired from the Air Force as a colonel in 2010 after serving as a combat analyst during initial operations in Iraq, directing an effects assessment survey team in the summer of 2003, and leading the Strategic Assessments cell, Multi-National Forces-Iraq, from 2007 to 2008. He holds a Master's Degree in Operations Research and a Doctorate in International Economics.